

Light-Weight On-Chip Monitoring Network for Dynamic Adaptation and Calibration

Pablo Ituero, *Member, IEEE*, Marisa López-Vallejo, *Member, IEEE*,
Miguel Ángel Sánchez Marcos, and Carlos Gómez Osuna

Abstract—Current nanometer technologies suffer within-die parameter uncertainties, varying workload conditions, aging, and temperature effects that cause a serious reduction on yield and performance. In this scenario, monitoring, calibration, and dynamic adaptation become essential, demanding systems with a collection of multi purpose monitors and exposing the need for light-weight monitoring networks. This paper presents a new monitoring network paradigm able to perform an early prioritization of the information. This is achieved by the introduction of a new hierarchy level, the threshing level. Targeting it, we propose a time-domain signaling scheme over a single-wire that minimizes the network switching activity as well as the routing requirements. To validate our approach, we make a thorough analysis of the architectural trade-offs and expose two complete monitoring systems that suppose an area improvement of 40% and a power reduction of three orders of magnitude compared to previous works.

Index Terms—Aging, complementary metal oxide semiconductor, dynamic adaptation, leakage, monitoring, on-chip, temperature.

I. INTRODUCTION

CURRENT nanometer technologies have allowed extraordinary integration densities in digital circuits. However, as technology scales down into the deep nanometer regime, design considerations for yield and reliability of operation have become critical. Transistors vary significantly from wafer to wafer and from die to die because of process uncertainties. Moreover, the device characteristics change over time as defects accumulate within the transistors due to activity and wearout.

The traditional way of fighting against these harmful effects was defining guard-bands which hid the possible variations using a corner-based design approach (typically such that 3σ of all manufactured circuits would not exceed corner values). This safe and inefficient way of designing does not apply any

more, since the area-performance-power budget is shrinking fast, and the process-induced variation and time-dependent shifts in transistor parameters are increasing rapidly [1].

Another battle front is at power densities and operating temperatures of the circuits, which continue to rise at an alarming rate. Dynamic power and thermal management (DPM and DTM) appeared as solutions to avoid spatial and time distributed hotspots sustaining current performance improvement trends. Moreover, faults due to failure mechanisms like negative-bias temperature instability (NBTI) and time-dependent dielectric breakdown (TDDB) present an exponential dependence on temperature. This translates into shorter circuit lifetimes, that should be considered by DPM/DTM policies to maintain the levels of reliability and life-expectancy that consumers have come to expect.

In this context, proactive approaches to reliability become a must. Designers should develop architectures adaptable to variations of all kinds, which rely heavily on information gathered from in situ monitoring circuits. Multiuse sensors that can monitor performance, degradation, temperature or power consumption as the circuit ages are critical [2]. However, allocating an arbitrarily large number of such monitors will not only create a significant area overhead, but routing the data from the sensors to a central processing unit will also pose a challenge [3].

Our paper focuses on the new adaptive techniques based on real-time monitoring. In particular, we target the network that connects the monitors to a controller. We propose a new ultra light-weight network architecture able to tackle with different kinds of monitors working at different rates. The model simplifies the calibration process, is compatible with all existing monitoring standards and reduces the packet traffic by data selectivity. The main contributions are the following:

- 1) Proposal of a new interconnection hierarchy level, the threshing level, in between the monitors and the traditional peripheral buses that applies data selectivity to reduce the amount of information that is sent to the central controller.
- 2) To cover the new interconnection level, we introduce a single-wire monitoring network based on a time-domain signaling scheme that significantly reduces both the switching activity over the wire and the power consumption of the network. This scheme allows a straightforward obtention of an ordered list of values from the maximum to the minimum.

TABLE I
MONITORS REVIEW

Magnitude	Cause	Adverse effect	Control action	Period	Density	Quantization
Temperature	Localized power consumption	Performance/Reliability loss	DVFS/Clock throttling	ms	hundreds/chip	~8 bits
Critical path	Timing uncertainties	Speed loss	DVFS	μs	hundreds/chip	~7 bits
Supply voltage	Supply network impedances	Voltage droop/Malfunction	State recovery	ns	few-tens/chip	~6 bits
Aging	Prolonged usage	Speed reduction/Malfunction	DVFS	ms	hundreds/chip	~4 bits

3) We apply this signaling scheme to achieve digitization resource sharing in time and space of time-to-digital monitors.

To validate the approach, we have performed a thorough analysis of the architectural trade-offs of the network. Also, we have conceived a complete monitoring system for a single and an eight-core chip, including temperature and aging/process variations monitors. Our system displays a reduced area, power and complexity when compared with previous works.

Concerning the organization of the paper, in the next section, we review the most significant previous works and develop a list of ideal features of a monitoring network. In section III, we establish our hierarchical architecture model that includes the new threshing level and we put forward our time domain signaling scheme that fits conveniently in the threshing level. Section IV addresses the low level implementation details. Sections V and VI deal with implementation issues and architectural trade-offs. The case studies are described in section VII, followed by the concluding remarks.

II. RATIONALE

An on-chip monitoring system is composed of a set monitors that provide an output dependent on the corresponding physical magnitude; an interconnection network that delivers the data from the monitors to a controller; and a control system that can be either centralized or distributed. This work focuses on the interconnection network and the main objective is to provide a network architecture with a high degree of simplicity and a minimum impact in area and power.

There is a number of important physical and electronic magnitudes that, due to their uncertainty and unpredictability, require a monitoring system to track them. Some of these magnitudes, such as temperature, have been already on-chip monitored for over 20 years, whereas others have started to gain attention more recently when process variations and aging have become more and more relevant. As a start point, we ground the analysis of the monitoring network with a study of the requirements of the most important of these magnitudes. In particular, we have gone through temperature [4]–[6], critical path [6], supply voltage [6] and aging monitors [7] and have analyzed the characteristics that will display some impact in the design of the monitoring network. Although this list is not exhaustive, the aim is to put forward the most representative ones and, especially, those that are more restrictive. The results of the review are exposed in Table I. As shown, the sampling frequency, quantization and density needs vary greatly depending on the type of monitor. In the case of supply voltage monitors, they require such a fast control action that

they are normally directly attached to a distributed control system that takes immediate action, thus they fall out of the reach of this work.

Concerning previous works targeting the interconnection net, since the research community has paid attention to the thermal-aware design for a longer time, it is this field that has gathered most of the previous proposals. Apart from many approaches that just employ point-to-point connections to reach the temperature sensors, the first innovative approach that we found in the literature is that by Székely et al [8]. This pioneer work established all the basis of thermal-aware electronic design from thermal simulation to thermal monitoring. They proposed to insert the thermal test circuitry into the boundary-scan architecture and compare all the temperatures to a maximum rating. This idea of connecting all the monitors through a one-wire chain emulating a global shift-register imposes a lower bound in the routing of the network and has been employed frequently in the literature. Recent standards used in state-of-the-art processors, such as the Platform Environment Control Interface (PECI) by Intel, also make use of single-wire interfaces with the monitors.

The first work completely dedicated to the topic is [9], expanded afterwards by [10]. These works exposed the problem and proposed a network architecture that targets multicore processors and supports priority-based data transfer and customized interfacing. They prove the benefits of the employment of a dedicated monitoring network and include information of the network latency under several configurations to deliver all monitor data to the central controller.

An interesting analysis on interconnection architectures for hierarchical monitoring was presented in [11]. Based on mesh-based NoC platform experiments, they conclude that physically separate networks provide flexible and energy-efficient transmission for monitoring communication, with guaranteed latency. In particular, hierarchical monitoring networks are the most appropriate solution on the chosen platform.

Considering the requirements imposed by the monitors, along with the all the previous proposals, we next present a list of key features that we consider a network of on-chip monitors must fulfill:

- 1) **Multi Purpose.** Given that the network is the only way to access the monitors, it has to accomplish two main functions: monitoring and calibration. The network must have the delay and latency characteristics imposed by the control policies, and also must provide with unique and differentiated access to each of the sensors to realize the calibration.

- 2) **Ultra Light Weight.** The network must suppose a small overhead in terms of area, reliability, latency, power and self-heating of the system. A solution that fulfills the system requirements will necessarily go through a trade-off between area —especially the routing of each sensor to a processing unit—, power —mainly dependent on the data frequency and the switching activity of the interconnection lines— and both the amount and delay of the information that eventually gets to the policy controller.
- 3) **Flexibility.** The flexibility to host as many different types of monitors as possible is another key feature that will allow the network to be implemented in a variety of systems. A consequence of this need for adaptability is the requirement for standard interfaces that not only go towards the monitor-ends but also cover the network-OS and the network-PCB interfaces.
- 4) **Priority.** A network involving different types of monitors deals with data at diverse priorities. For instance, a quick action must be taken at general-emergencies, such as a big voltage supply droop or the surpassing of the safe limit in the junction temperature and this must be compatible with timely deliver of all the fixed rate information.
- 5) **Hierarchy.** A flat hierarchy network with all the monitors accessing the controller, would add a lot of complexity to the monitors, set hurdles to establish priorities and limit the number of nodes. Therefore a network with hierarchy levels is desirable.
- 6) **Dynamic Sensor-Selection.** Yet another feature is the dynamic sensor-selection to avoid collecting data from those sensors that will not provide useful information, as proposed in [4]. Ideally, the network should prevent the sensor from working whenever its information is not used.

Bearing in mind all these ideas, in the next section we propose a modification to the monitoring network architecture with the aim of reducing the network traffic and, thus, its bandwidth requirements.

III. NETWORK PROPOSALS

An important question that arises when dealing with the potentially enormous amount of data coming from an elevated number of on-chip monitors at high rates is if the central controller needs all that information. It is not a matter of compression —since the monitors are already quantized by the minimum number of bits to achieve the precision required by the control policy—, but of selectivity.

In most cases, the controller will just use a small set of values if not just one of the boundary values, either the biggest or the smallest [12], [13]. For instance, in a DVFS control system for a single core chip, just the delay from the slowest critical path and the highest temperature on-chip are necessary. Therefore, we propose to provide the monitor network with a certain intelligence that allows it to filter and select the information that eventually reaches the controller.

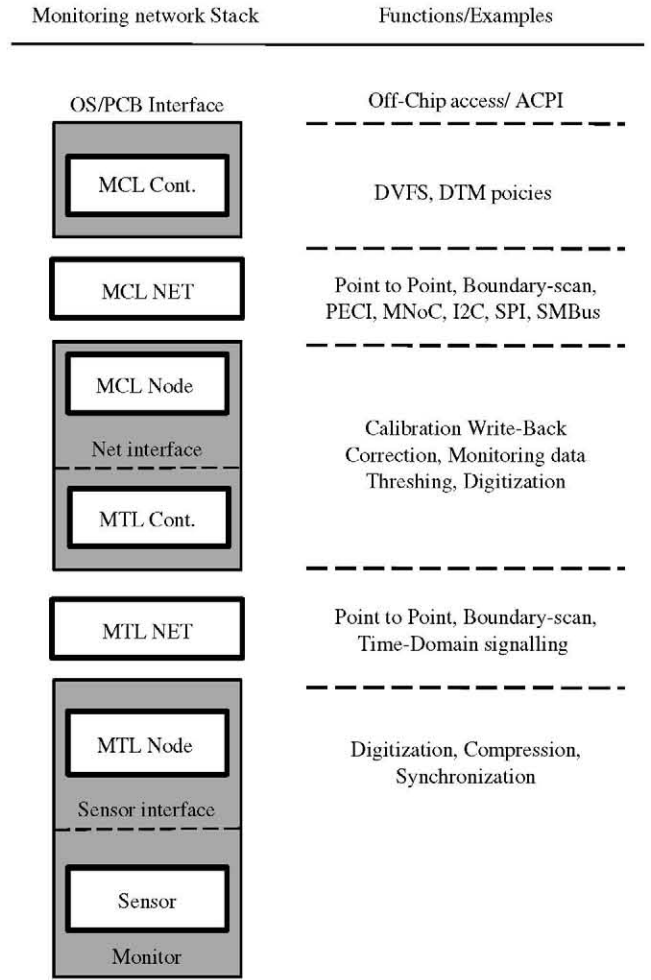


Fig. 1. Monitoring network stack proposal.

A. MTL

Specifically, we propose the architecture illustrated in figure 1 in which a new level of interconnection between the monitors and the traditional monitoring network is introduced. This new level gathers the data of each and every monitor on its subnet and, performing some elemental operations, decides which piece of data advances to the higher level in the hierarchy, for this reason, we refer to this level as Monitor Threshing Level (MTL). Besides, this level keeps the calibration write-back information of its monitors, freeing the central controller from this issue. In contrast, we refer to the upper level as Monitor Collecting Level (MCL), in this case the network delivers all the packets outgoing the MTL to the central controller.

The MTL must be able to forward to the next level the data of each monitor, uniquely identifying the source of the information. This is so, because in the calibration process it is possible that each monitor is required for evaluation. The logic operations expected to occur at the MTL entail calculating maximum and minimum values and averaging.

Next, we propose a new network architecture that targets the MTL employing time-domain signaling and digitization resource sharing.

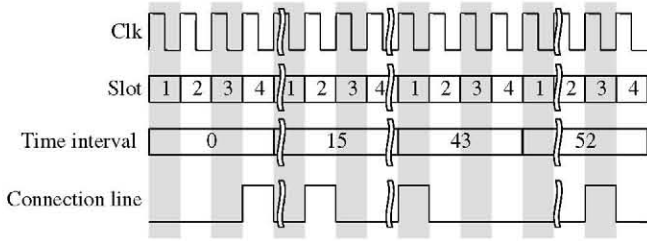


Fig. 2. Time-domain signaling chronogram for a 4-sensor network employing 2-bit data.

B. Time Domain Signaling

A network topology targeting the MTL must be able to collect all the monitor data employing as little power and as fewer interconnections as possible. The boundary-scan architecture [14] establishes a lower bound in the connection part, but entails an elevated activity in the network. Since the monitors are distributed across the entire chip, the power cost of charging and discharging the interconnection capacitance cannot be neglected. In order to reduce the activity in the network, we propose a signaling scheme that just requires a single pulse to send the digital data from each monitor to the controller without any loss of information.

For an n -monitor network in which all the monitors quantize the information with the same number of bits q , we propose to divide the time into 2^q intervals and each interval into n slots, so that each slot corresponds to a monitor and each interval to a possible value of the digital word. For example, figure 2 illustrates a four-monitor network in which monitors 1-4 send the values 43, 15, 52 and 0, respectively. Now, if we analyze the number of transitions produced in the interconnection line, we find that the upper bound for an n -monitor network is $2n$, which means a significant reduction compared to other single-wire serial-bit transmission schemes.

To make the conversion from the time domain to the digital domain, a counter at the controller that keeps the sequence of time intervals is enough. Another counter is able to identify the monitor that produced the pulse. Concerning the network intelligence, certain operations come very naturally with this type of signaling. Such is the case of calculating the maximum and the minimum of all the data or ordering of the information as a way of providing priority. For instance, if the network just needs the maximum measurement, the coding can be realized in reverse order, such that the higher values get shorter times; in this way the maximum value arrives first and then the rest of the system can be stopped.

Next we describe a case of use of the proposed time-domain signaling that applied to a certain type of monitors supposes a better power and area reduction by the sharing of digitization resources.

C. Digitization Resource Sharing

Each monitor is composed of two main parts. First, a sensing block that yields an analog signal with a well known dependency on the magnitude we want to measure and a digitization block. When the analog signal is a voltage or a current, the digitization block is an ADC and when the analog

signal is a pulse with a varying width or other time dependent signal, the digitization block is a time-to-digital converter. The digitization block is the part that normally occupies more area and consumes more power [5].

Let us gain some insight on the monitors that comply with the latter case where the analog signal is time dependent. Every critical-path monitor has an implicit time-to-digital conversion [6] and many of the aging monitors that have been proposed are based upon the variations of the delay of a certain path. A whole generation of temperature sensors based upon time-to-digital converters have appeared in the last few years imposing a new paradigm because of their reduced power consumption and area. For example the sensors in [5] make use of the leakage current thermal dependencies to produce the pulse.

Interestingly, the analog signals provided by this type of sensors can be bypassed to our network without the digitization step, by means of a synchronization module that catches the signal transition and inserts a pulse in the network in the next available time slot. In this way, all the digitization blocks distributed across the network at each monitor are replaced by much simpler synchronization modules, and the digitization of all the monitors is performed at the controller. This represents an important saving in area and power consumption. Note that this scheme, at most, produces an error of a counting period which compares to the one produced by the standard counter-based time-to-digital conversion.

D. Calibration

On-chip monitors require a calibration phase to improve their accuracy and linearity. During this costly process, carried out under controlled conditions, the monitors are read and the write-back information is stored. This information usually consists of a set of linear correction parameters and serves for one or several monitors. In our hierarchical network with prioritized routing, it is mandatory that the calibration write-back information is stored in the lower levels of the structure. In this way the correction can be applied as soon as the data is produced.

In our architecture, for all the calibration conditions, all the monitors are read and their corresponding MTL controller stores all the values. Next, the MCL retrieves all these values for each of the calibration corners. After generating the necessary write-back information, it is conveniently stored in the MTL controller though the MCL network. The MTL controllers store the write-back information and have a dedicated and shared datapath that applies the correction of the incoming data.

IV. DETAILED NETWORK DESCRIPTION

In this section we provide low-level hardware details of the implementation of the proposed network scheme including the different modules that conform it. The network architecture is based on the following principles:

- 1) The network consists on one central controller, or front-end, and several monitor nodes, or back-ends, which share one physical data channel.

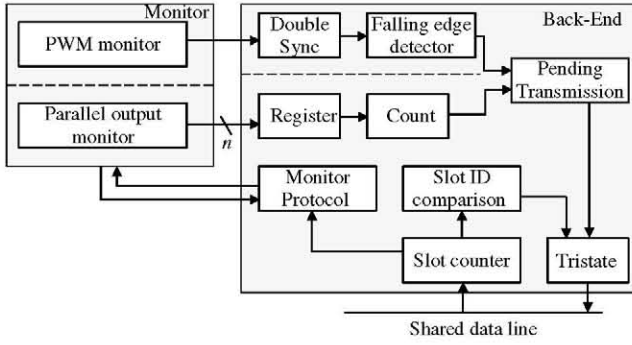


Fig. 3. Implementation of the back-end module.

- 2) Data channel access is based on a time division mechanism, with fixed slots for each of the nodes to avoid data collision. Each node knows its slot on the multiplexing scheme and all nodes are synchronized to avoid collision, no negotiation or arbitration mechanism is present to reduce complexity to the network.
- 3) Apart from the data channel, one clock and reset lines are available for both masters and slaves within the network.
- 4) Each back-end is connected to a monitor.
- 5) Each back-end is responsible for the generation of the request/acknowledge interface towards its sensor.
- 6) The central controller uses slot 0 in the time division mechanism to send the control instructions to the rest of the nodes.
- 7) The central controller converts the answer of each of the nodes into a digital time measurement.

Concerning the operation, all back-ends and the front-end share the same line on which each node has been assigned a one cycle slot to send its data. This time-slot is hard-coded at implementation time. The front-end of the network sends the control instructions employing slot 0 and then waits for the back-ends data to arrive through the serial line. Each back-end node interfaces with its corresponding monitor through a two line request/acknowledge control interface plus a data line in which the sensor will drive either the PWM signal or the parallel digital signal. In the case of PWM monitors, each back-end synchronizes the detected end of the pulse with the next available time-slot. In the case of monitors with a parallel interface, the back-end transforms the information into a PWM signal by means of a counter. The end of the count is synchronized with the next available time-slot.

A. Network Back-End

This module realizes three main functions. First, it has to listen to the network to determine when the a monitoring round starts. Second, it interfaces the sensor, informing it when to start the measurement, and receiving the sensor outgoing data. And third, it has to synchronize the monitor data so it is asserted at the corresponding time slot. This module is depicted in figure 3 for both PWM and parallel output monitors.

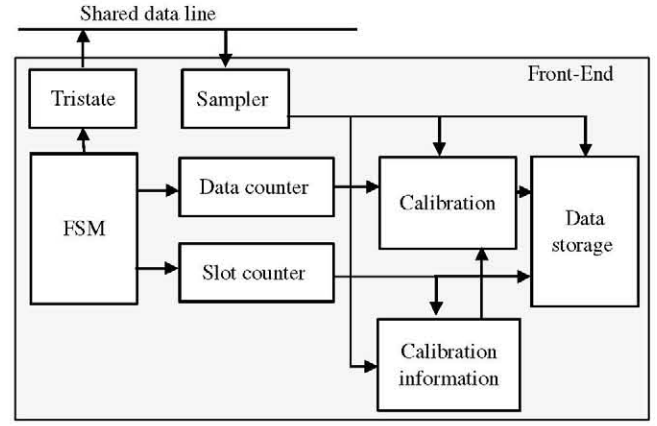


Fig. 4. Implementation of the front-end module.

TABLE II
MINIMUM RELIABLE TIME SLOTS FOR DIFFERENT WIRE LENGTHS

Width/ Length	500 μ m	1 mm	2 mm	4 mm	8 mm	16 mm
300 nm	500 ps	700 ps	1.4 ns	2.7 ns	5.3 ns	12 ns
400 nm	600 ps	1.0 ns	1.7 ns	3.2 ns	6.3 ns	14 ns
500 nm	600 ps	1.1 ps	1.9 ns	3.6 ns	7.2 ns	16 ns

When the control instructions are sensed, the back-end node becomes active, sending the request signal to the sensor, waiting for its acknowledge. In the case of a PWM sensor, a falling-edge detector registers the end of the pulse from the sensor. In the case of a sensor that provides a parallel digital signal, the back-end stores the signal and activates the counter that converts this signal into a varying-width pulse.

After this, the back-end node waits for its turn in the serial line and sends a 1 cycle-long high pulse through it to indicate the end of the measurement. Slot control is achieved by an internal counter which is synchronized with all other nodes on reset.

B. Network Front-End

The front-end controller either produces a periodic start pulse at an established sampling frequency or it can also act in an on-demand mode in which it waits for an external signal activation to assert the start pulse. In either mode, the front-end employs slot 0 to send the control instructions. Subsequently, it waits until all sensors in the network have sent their end-of-measurement pulse. Once the front-end has received pulses from all the nodes, it goes back to its idle state.

Concerning the digitization process, each time a pulse is received on the shared line, the outgoing digital value from the counter, that indicates the number of cycles from the beginning of the readout process, undergoes a linear transformation according to the stored calibration data. This corrected value along with the monitor identification, extracted from the time slot, are stored in a data storage unit.

In order to implement these functionalities, the block, depicted in figure 4 includes the following internal structures:

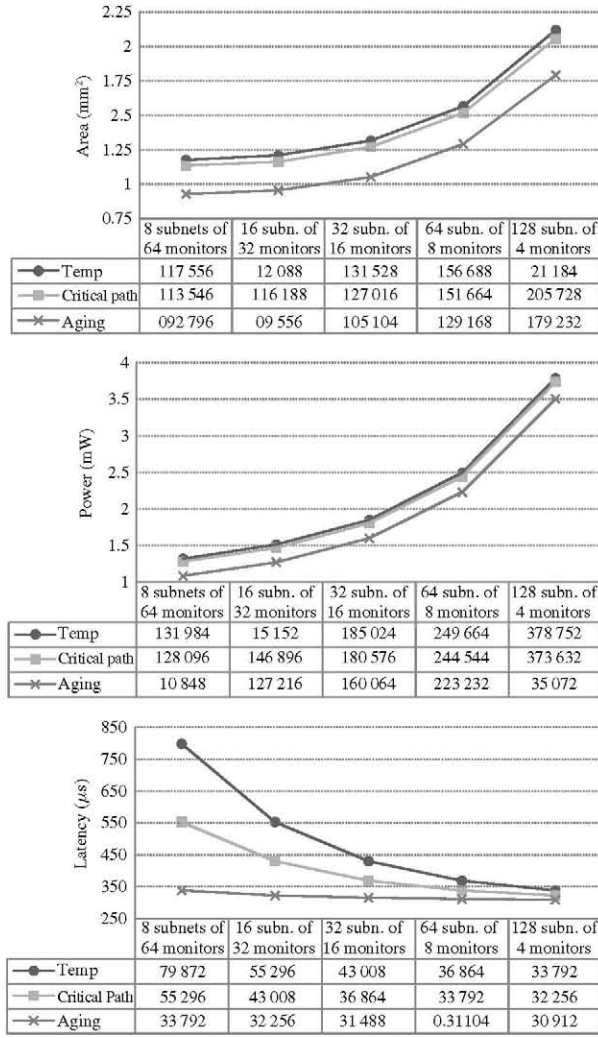


Fig. 5. Architectural trade-off analysis for a 512-node network of temperature, critical path, and aging monitors.

- 1) An FSM to distinguish between the different operating modes.
- 2) A slot counter in order to synchronize with all back-end nodes.
- 3) A bank of counters, to generate the time to digital conversion measurements.
- 4) The calibration logic that performs the transformation of the measurements.
- 5) A data storage unit containing the current measurements.
- 6) A memory containing the calibration information.

V. SHARED LINE IMPLEMENTATION ISSUES

The reliability and the robustness of the architecture greatly depend on sets of single pulses traveling along a shared transmission line. This can appear as a risky approach, due to several sources of problems, such as variability and noise issues, however, the low bandwidth requirements of monitoring networks allow to reduce the working frequency down to a completely safe value while delivering all the necessary data on time.

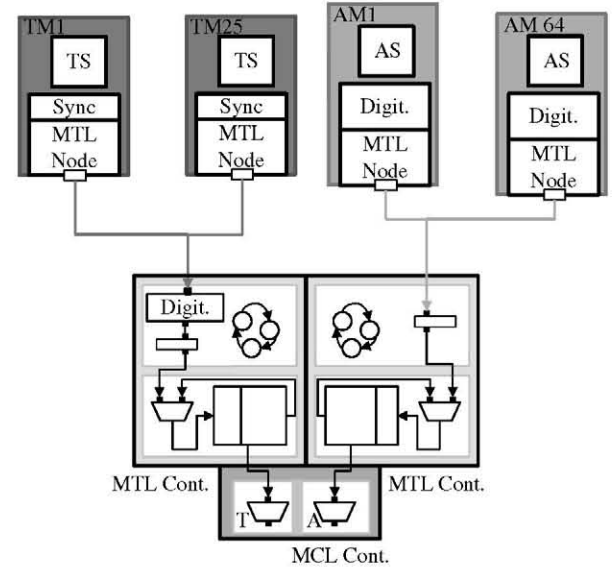


Fig. 6. Block-level description of the single-core monitoring system.

In order to determine the working frequency that provides error-free functioning for a target line length we have performed extensive Monte Carlo simulations varying the parameters that affect the quality of the traveling pulse. For this study we have employed a 6-metal 90 nm technology from UMC. We have covered process, V_{DD} ($\pm 10\%$ uniform distribution) and thermal (industrial range -40°C to 85°C) variations. We have altered the position of the pulse transmitter so that the differences in transmission time are accounted for. And, finally, we have considered several crosstalk scenarios. First, the type of signaling was also taken into consideration, however, for the sake of simplicity, we did not employ any special signaling eventually.

The sources of crosstalk are completely determined by the layout of the circuit that the network monitors. We have assumed that the line is implemented in middle metallization layers, which are more likely to be able to accommodate some extra wiring, since lower levels are employed for the implementation of the standard cells and upper levels are normally reserved for system levels signals such as V_{DD} , ground and clocks. In particular we have carried out our experiments with a metal 3 layer (out of 6). On this layer, most capacitance is to other signal lines. In the worst scenario, which has been considered in the analysis, another data line goes in parallel so that the parasitic capacitance is maximum. In a more realistic scenario, our transmission line could cross the bit lines of a datapath. The values on the bit lines are highly correlated and may all switch simultaneously in the same direction [15]. This case has also been accounted for.

A data line with several back-ends distributed at different locations has the extra problem of a variable transmission time from the back-ends to the listening front end. This effect also takes into account the action of the clock skew at the back-ends. The net clock is transmitted from the front-end so that all the back-ends receive a slightly delayed version of it. When a back-end transmits a pulse, it arrives at the front-end with

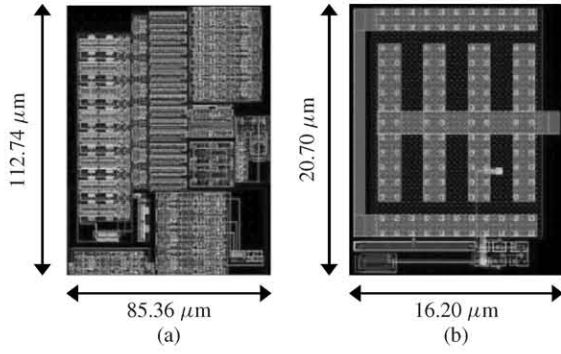


Fig. 7. Layout detail of the aging/process and temperature monitors. (a) Aging/process monitor. (b) Temperature monitor.

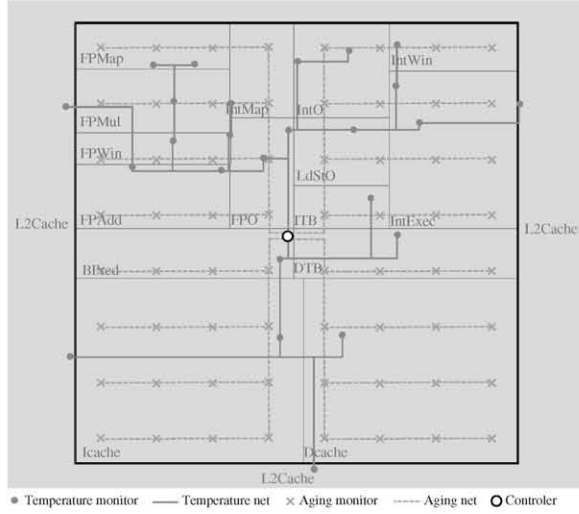


Fig. 8. Shared interconnection lines routing for the single-core monitoring system.

a double delay since it was produced employing a skewed clock and then it took some time to travel through the data line. These problems were simulated by placing three backends, one at each end and another in the middle. The results of the analysis for different interconnection line lengths and widths are shown in Table II. As shown, the capacitance of the line is the most restrictive factor and it must be kept as thin as possible. Due to spacing limitations of upper metal layers, normally wider and thus entailing a bigger capacitance, the shared interconnection line must be kept as low as possible.

Note that one controller can share the same digitization resources among different connecting lines. In this case we say that all the monitors that employ the same count are connected by a logical network. For example a controller could have two interconnection lines employing the same counter such that the first line occupies a portion of the time slots in a time interval and the second line occupies the rest. The location of the controller is, thus, an important issue to consider because it can facilitate the partition of logical networks into different physical connecting lines.

VI. ARCHITECTURAL TRADE-OFFS

In order to analyze the optimum size of the MTL subnetworks with our time domain signaling, we have performed an

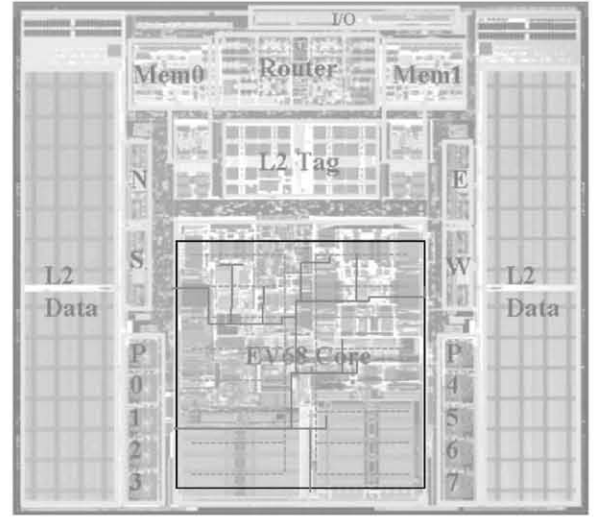


Fig. 9. Mock-up of the layout of the single-core monitoring system.

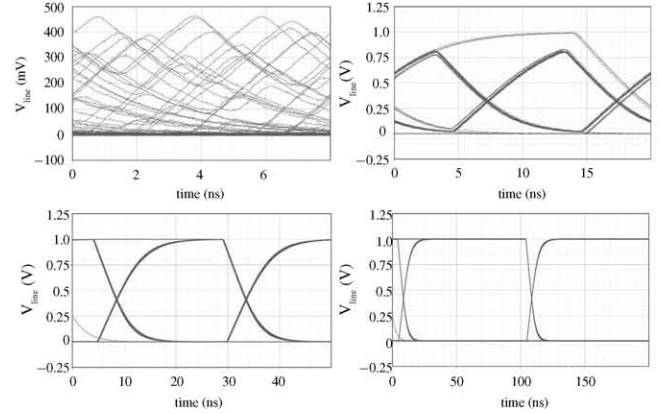


Fig. 10. Single Core monitoring system.

implementation analysis for a global network of 512 monitors varying the number of monitors in the MTL subnetworks from 4 to 64. For the MCL network, we have employed the I^2C standard. This analysis has been realized for three types of monitors: temperature, aging and critical path which require different quantization schemes and sampling frequencies as displayed in Table I. The variables considered in the study are the area and power overheads of the network along with the latency of the network to deliver the data from all the monitors to the controller. The designs have been synthesized on a 90 nm technology using a 100 ns clock.

Figure 5 shows the results. As a general tendency, the area and the power increase as the number of subnetworks rises, this is basically because of the growth in the number of I^2C nodes, however the latency decreases as the time intervals become shorter. It is important to consider that the bigger the number of monitors inside a subnetwork, the longer the single wire connection is, and therefore, the more restrictive the network clock is. In the case of the critical path monitor, the long latency, imposed by the I^2C standard, makes it impossible to fulfill the sampling rate requirement which is in the order of microseconds, however a threshing policy could mitigate this problem.

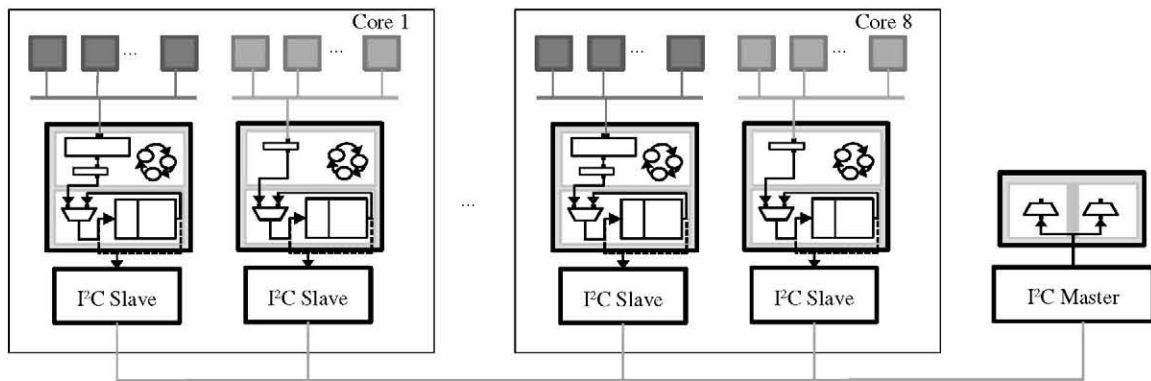


Fig. 11. Eight core monitoring system.

VII. CASE STUDY

To validate the advantages of our proposals, in a first experiment we have implemented a complete monitoring system for a single-core Alpha 21364 processor. The features of this processor were presented in [16], it was fabricated in a $0.18\ \mu\text{m}$ process, entailing an area of $21.1 \times 18.8\ \text{mm}^2$ and containing 152M transistors. All of our experimental work was carried out in the 90 nm technology node, therefore, in order to have sensible values for on-chip distances, we have applied the $1/\sqrt{2}$ rule twice in each dimension of the original chip corresponding to two technology generation steps. The system includes temperature and aging/process variations monitors, a high-level description is shown in figure 6. To implement the network modules, we have employed a low power standard cell library from Faraday targeting UMC's 90 nm process and the numeric results come from the synthesis simulation under typical conditions. The monitors have been implemented in full-custom and validated through parasitic-extracted analog simulations.

The aging/process variations data are fetched every 1 ms and transmitted with our time-domain signaling to the MTL based frontend. We have adapted the critical-path monitor in [17] to our technology, so that a 15-inverter thermometer code is converted to obtain a 4-bit quantization signal. Figure 7(a) shows the layout of the monitor which entails an area of $9623\ \mu\text{m}^2$. The modeling of the process variations was performed with the technology parameters probability distributions provided by the technology foundry. For the aging modeling, we used the methodology proposed in [18], namely the negative bias temperature instability (NBTI) is simulated as a logarithmic decay of V_{th} with time. Concerning the placement and allocation of the monitors, based on [13], that targets the same processor, we have allocated 64 monitors uniformly distributed across the core.

For the thermal monitoring we have adapted to our technology the temperature sensor described in [5]. Figure 7(b) shows the layout of the monitor which supposes an area of $335\ \mu\text{m}^2$. This reduced area is due to the fact that, in this network, we employ the proposed digitization resource sharing, thus just the sensing part is included. In order to adapt the monitor to our timing requirements, we have added a MIM capacitor that expands the width of the varying-width signal so that the

monitor provides a pulse with a varying width up to $750\ \mu\text{s}$ for a range of temperatures of $40 - 90\ ^\circ\text{C}$ with a sampling rate of 1 KHz. As shown this capacitor takes up most of the space of the monitor. A complete analysis of the thermal-aware control policies for this microprocessor was described in [12]. As far as the placement and allocation of the monitors is concerned, for the Alpha 21364 processor, Memik et al. in [3] considered several temperature sensor allocation schemes. For our network, we use the scheme referred to as "hybrid-allocation" that allocates 25 monitors.

In order to find the optimal location of the controller, we first completely routed each of the networks connecting all the monitors. We modified the output of the Lee Algorithm [19] so that a line of minimum capacitance was achieved. This algorithm minimizes the delay between one node and the rest, which does not necessarily imply that it takes the minimum capacitance connection. However, the modifications were very slight, in the case of the temperature network just one segment was moved, whereas no change was necessary in the aging network. Once the routing was performed, we decided to divide the logical temperature network into two physical connections and the aging one into four. Figure 8 shows the routing of both networks. The detail of the size of the network compared to the size of the whole chip is displayed in figure 9. The die photo of the Compaq Alpha 21364 has been obtained from CPU Info Center website.

The most restrictive transmission line of the six resulting physical connections is the upper temperature line, since it is the one that entails a bigger number of monitors and a bigger capacitance. In order to determine the characteristics of the traveling signal in the network, we performed eye diagram simulations employing different time slots, as shown in figure 10. We have employed a base of metal three for the interconnection, but we have introduced a number of random metal changes to simulate the adaptation to underlying architectures. Furthermore we have introduced noise sources simulating clock and data lines cross-talk. As shown, beyond 10 ns the eye opening allows for a completely reliable transmission of the signal.

In the final design, both networks employ a 100 ns slot, which yields $6.40\ \mu\text{s}$ and $2.50\ \mu\text{s}$ intervals for the aging/process variations and the temperature networks, respectively. Note that we have used a rather slow frequency,

TABLE III
SYNTHESIS RESULTS OF THE IMPLEMENTED MONITORING NETWORKS

	Single-Core				Eight-Core					
	T Node	A Node	Controller	Total	T Net (X1)	T Net (X8)	A Net (X1)	A Net (X8)	I ² C Net	Total
Area (μm^2)	1103	1199	166547	228000	33400	267200	41300	330400	9840	1430000
Power (μW)	1.34	1.35	56.3	174.5	66.7	533.6	117.5	940	289.6	1763.2

compared with other on-chip interconnections, and however it suffices our bandwidth requirements. In the case of the aging monitors, all the information is transmitted in 102 μs , in the temperature network the latency is 640 μs .

In a second experiment, we addressed an eight-core chip, based on the Alpha 21364 processor. Figure 11 describes the architecture, for each core we have allocated the same monitors as in the previous experiment, but in this case we see that we have two levels in the network hierarchy. The intermediate nodes are the masters of our MTL and slaves of an I²C serial network. The general dynamic management controller is the master of the I²C connection. The latency of this network is 844 μs .

Table III shows the results of the synthesis characterization. In the eight-core architecture, for the power simulation the network delivers all the monitors information to the controller over the I²C infrastructure, thus, no threshing is performed. Let us compare these results with the monitor interconnect subsystem exposed in [9] where also a Faraday's 90 nm standard cell library is used and the sampling frequency is also 2 ms. In this work they employ 192 temperature monitors spread across an eight-core architecture, using 16 routers and obtain an area of 0.819 mm^2 and a power consumption of 244 mW. Taking just the 200-monitor temperature section our architecture along with the I²C infrastructure, we get an area of 0.432 mm^2 , 40% less, and a power consumption of 0.36 mW, which suppose an improvement of three orders of magnitude. This is coherent with the fact that our system employs a much lower frequency than [9].

VIII. CONCLUSION

In the late CMOS era, the challenges imposed by reliability, aging and thermal issues, among others, have highlighted the need for monitoring, calibration and dynamic adaptation. This article has approached the interconnection infrastructure that delivers data from a set of distributed monitors to a central controller. We have proposed a new interconnection hierarchy that allows information selectivity so that just the necessary information arrives to the policies controller and we have presented a time-domain signaling scheme over a single wire that on the one hand importantly reduces the switching activity and the power consumption and on the other, naturally eases the ordering of the information and the calculation of the maximum and minimum values, basic tasks for priority-based selectivity. We have analyzed the architectural trade-offs and presented two prototypes of complete monitoring systems that have proved to significantly overcome previous works in terms of area and, specially, power consumption.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and H. A. Cerqueira for his support on the aging/process monitor.

REFERENCES

- [1] M. Alam, "Reliability- and process-variation aware design of integrated circuits," *Microelectron. Rel.*, vol. 48, nos. 8–9, pp. 1114–1122, 2008.
- [2] D. Sylvester, D. Blaauw, and E. Karl, "Elastic: An adaptive self-healing architecture for unpredictable silicon," *IEEE Des. Test Comput.*, vol. 23, no. 6, pp. 484–490, Jun. 2006.
- [3] S. Memik, R. Mukherjee, M. Ni, and J. Long, "Optimizing thermal sensor allocation for microprocessors," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 27, no. 3, pp. 516–527, Mar. 2008.
- [4] J. Long, S. Memik, G. Memik, and R. Mukherjee, "Thermal monitoring mechanisms for chip multiprocessors," *ACM Trans. Archit. Code Opt.*, vol. 5, no. 2, pp. 1–33, 2008.
- [5] P. Ituero, J. Ayala, and M. Lopez-Vallejo, "A nanowatt smart temperature sensor for dynamic thermal management," *IEEE Sens. J.*, vol. 8, no. 12, pp. 2036–2043, Dec. 2008.
- [6] A. Drake, *Adaptive Techniques for Dynamic Processor Optimization*. New York: Springer-Verlag, 2008.
- [7] P. Singh, C. Zhuo, E. Karl, D. Blaauw, and D. Sylvester, "Sensor-driven reliability and wearout management," *IEEE Des. Test Comput.*, vol. 26, no. 6, pp. 40–49, Nov.–Dec. 2009.
- [8] V. Szekely, M. Rencz, and B. Courtois, "Tracing the thermal behavior of ICs," *IEEE Des. Test Comput.*, vol. 15, no. 2, pp. 14–21, Apr.–Jun. 1998.
- [9] S. Madduri, R. Vadlamani, W. Burleson, and R. Tessier, "A monitor interconnect and support subsystem for multicore processors," in *Proc. Des. Automat. Test Eur. Conf. Exhibit.*, 2009, pp. 761–766.
- [10] J. Zhao, S. Madduri, R. Vadlamani, W. Burleson, and R. Tessier, "A dedicated monitoring infrastructure for multicore processors," *IEEE Very Large Scale Integr. Syst. Trans.*, vol. 19, no. 6, pp. 1011–1022, Jun. 2011.
- [11] L. Guang, E. Nigussie, J. Isoaho, P. Rantala, and H. Tenhunen, "Interconnection alternatives for hierarchical monitoring communication in parallel socs," *Microprocess. Microsyst.*, vol. 34, no. 5, pp. 118–128, Aug. 2010.
- [12] K. Skadron, M. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan, "Temperature-aware microarchitecture: Modeling and implementation," *ACM Trans. Archit. Code Opt.*, vol. 1, no. 1, pp. 94–125, Mar. 2004.
- [13] A. Tiwari, W. Anderson, J. Torrellas, "Facelift: Hiding and slowing down aging in multicores," in *Proc. IEEE/ACM 41st Annu. Int. Symp. Microarchit.*, Lake Como, Italy, 2008, pp. 129–140.
- [14] V. Szekely, M. Rencz, S. Török, C. Márta, and L. Lipták-Fegő, "Cmos temperature sensors and built-in test circuitry for thermal testing of ics," *Sens. Actuat. A: Phys.*, vol. 71, nos. 1–2, pp. 10–18, 1998.
- [15] W. Dally and J. Poulton, *Digital Systems Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [16] A. Jain, W. Anderson, T. Benninghoff, D. Berucci, M. Braganza, J. Burnetie, T. Chang, J. Eble, R. Faber, O. Gowda, J. Grodstein, G. Hess, J. Kowaleski, A. Kumar, B. Miller, R. Mueller, P. Paul, J. Pickholtz, S. Russell, M. Shen, T. Truex, A. Vardharajan, D. Xanthopoulos, and T. Zou, "A 1.2 GHz alpha microprocessor with 44.8 GB/s chip pin bandwidth," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, 2001, pp. 240–241.
- [17] A. Drake, R. Senger, H. Deogun, G. Carpenter, S. Ghiasi, T. Nguyen, N. James, M. Floyd, and V. Pokala, "A distributed critical-path timing monitor for a 65 nm high-performance microprocessor," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, San Francisco, CA, 2007, pp. 398–399.

- [18] A. C. Cabe, Z. Qi, S. Wooters, T. N. Blalock, and M. R. Stan, "Small embeddable NBTI sensors (SENS) for tracking on-chip performance decay," in *Proc. Quality Electron. Des.* San Jose, CA, 2009, pp. 1–6.
- [19] C. Y. Lee, "An algorithm for path connections and its applications," *Electron. Comput. IRE Trans.*, vol. 10, no. 3, pp. 346–365, Sep. 1961.



Pablo Ituero (M'10) received the M.S. degree in telecommunications with a major in electronics from the Universidad Politécnica de Madrid, Madrid, Spain, in 2005, and the M.S. degree in electrical engineering, specializing in system-on-a-chip designs, from the Royal Institute of Technology, Stockholm, Sweden, in 2005. He is a Ph.D. candidate with the Department of Electronic Engineering, Universidad Politécnica de Madrid.

He is an Assistant Professor with the Department of Electronic Engineering, Universidad Politécnica de Madrid. His current research interests include high-performance processor architectures and low-power and process, voltage, temperature-aware electronic designs.



Marisa López-Vallejo (M'00) received the M.S. and Ph.D. degrees from the Universidad Politécnica de Madrid, Madrid, Spain, in 1993 and 1999, respectively.

She is an Associate Professor with the Department of Electronic Engineering, Universidad Politécnica de Madrid. She was with the Lucent Technologies, Bell Laboratories, Murray Hill, NJ, as a Technical Staff Member. Her current research interests include low-power, process voltage, temperature-aware designs, computer-aided diagnostic methods and tools, and application-specific high-performance programmable architectures.



Miguel Ángel Sánchez Marcos received the M.S. degree in telecommunications with a major in electronics from the Universidad Politécnica de Madrid, Madrid, Spain, in 2003. He is a Ph.D. candidate with the Department of Electronic Engineering, Universidad Politécnica de Madrid.

His current research interests include embedded systems and high-performance digital signal processing for field-programmable gate arrays.



Carlos Gómez Osuna received the M.Sc. degree in telecommunications engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2004.

He is currently a Systems Engineer for the Spanish Ministry of the Interior and a Freelance Researcher, UPM. His current research interests include real time debug infrastructures and low power digital designs and verification.